Astronomy & Astrophysics manuscript no. rn4 February 19, 2004
(DOI: will be inserted by hand later)

On the detection of pure sine waves embedded in a spectrum of stochastically excited p modes

O.Moreira[1], T.Appourchaux[1⋆], G.Berthomieu[2], and T.Toutain[2]

[1] European Space Agency, P. O. Box 299, 2200AG Noordwijk, The Netherlands
[2] Département Cassini, UMR CNRS 6529, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France

Received .../ accepted ...

Abstract. The modes identification and fitting of solar oscillations provide an observational mean to derive the physical properties of Sun's interior. For the future asteroseismology space missions, the application of the experience gained with solar data analysis, as well as automatic fitting procedures, have been applied to artificially generated time series. In this paper, we have devised a new technique for detecting long-lived modes embedded in a common p-mode spectrum. This technique has been applied to synthetic time series of the evolved solar-like star HD57006. It has also been validated using Monte-Carlo simulations. The impact of how the long-lived modes are excited on the detection level is also discussed.

Key words. Stars: oscillation – Methods: statistical

1. Introduction

Observed p-mode spectra have been used for more than two decades to study the internal structure of the Sun in great details. However, because the major contribution of the p modes is in the outer layers of the Sun and only the $l = 0$ modes reach the core, the inversion techniques based on p modes frequencies requires great accuracy to determine the core structure[1]. In contrast to the p modes, the g modes are confined to the radiative zone and have an energy maximum located near the center of the Sun. Therefore, the inversions of solar structure using g-mode frequencies do not require such a high frequency precision.

Seeking for g modes in solar data has always been the challenge of helioseismology. Statistical methods and pattern recognition techniques have been used aiming to detect g modes in the data collected by SOHO and ground-based instrument (Appourchaux et al., 2000; Gabriel et al., 2002; Turck-Chièze et al, 2004). Although, the detection of the g modes has not been successful, space missions such as MOST and COROT will provide us, in the near future, with measurements of light fluctuations from other stars with enough precision to detect and identify modes of oscillation of any kind, p and g modes alike. The knowledge achieved on the identification and tagging of the normal modes from full-disk observations of the solar global oscillations will be a great aid for performing the mode identification task on solar-like stars [2]

Aiming to prepare for the future space mission, the Data Analysis Team of the Seismology Working Group of COROT has been performing Hare-and-Hound (H&H) exercises. These kind of exercises has two steps. In the first step a team generates theoretical mode frequencies and synthetic time series. In the second step another team analyzes the time series, performs the mode identification, peak fitting and frequencies inversion without having access to any other information but the known characteristics of the star.

In the framework of the third H&H exercise, a time series of the star HD57006 was generated by one of the authors (TT) with frequencies provided by another author (GB) and the modes identification and fitting by the other two authors (OM and TA). The star HD57006 has been presented (Appourchaux et al., 2003) as an evolved star with 1.6 $M_\odot$ which has mixed modes of oscillation (modes that have a p-mode character in outer layers of the star and a g-mode character in the star core). Besides the problems related with modes identification due to the complexity of the oscillation spectrum predicted for this star, this time series also had, bellow 500 μHz, prominent peaks that correspond to long-lived modes amongst the short-lived p-mode spectrum. In this paper, we will briefly describe the statistical tests used in the detection techniques of long-lived modes or g modes (Appourchaux et

⋆ *Address as of April 1st, 2004*: Institut d'Astrophysique Spatiale, Batiment 121, 91405, Orsay Cedex, France

[1] It requires a frequency precision of the order of 10 nHz for a 5-minute p mode

[2] type of stars as defined by Appourchaux (2003)

al. 2000, and Gabriel et al. 2002). Next we will introduce a technique for detecting these long-lived modes when they are embedded in short-lived modes. The validation of the new detection technique will be done through a Monte-Carlo analysis and through an evaluation of the result on the synthetic time series of HD57006.

2. Long-lived modes detection

2.1. Detection of long-lived modes alone

A signal that has no detectable intrinsic width in the power spectrum (such as a pure sine wave or a mode with a lifetime longer than the observation duration) can be detected in a white noise signal by testing the following hypotheses:

H0: A prominent observed peak within the power spectrum might be merely due to noise (pure noise observed)

H1: A sine wave signal embedded in white noise might be detectable above a given significance power level (pure noise observed with a deterministic signal)

The statistical tests based on H0 consist in determining a significance level for which the peaks have a low probability of being due to noise. It is assumed that all power bins are independent identically distributed (i.i.d) random variables. The H0 hypothesis does not check for the presence of signal, this is the 'rôle' of the H1 hypothesis. The test based on the H1 hypothesis gives the probability that a sine wave of a given deterministic amplitude can be detected reaching a pre-determined significance level in the power spectrum. It is applicable when the characteristics of the embedded sine wave are known. These two kind of tests do not contradict each other but test two different assumptions.

2.1.1. Statistics of pure noise

The power spectrum of a pure noise signal made from full-disk integrated instrument has a χ^2 statistical distribution with 2 d.o.f. If $x_r(\nu_k)$ and $y_i(\nu_k)$ are normalized normal distributions, representing respectively the real and the imaginary part of the noise Fourier spectra, then the power spectrum can be described as:

$$P(\nu_k) = x_r^2(\nu_k) + y_i^2(\nu_k) \tag{1}$$

Which means if $z = x^2 + y^2$ is an observed value in the power spectrum then its probability distribution F is given by:

$$G(z) = 1 - e^{-\frac{z}{2}} \tag{2}$$

If we normalize we respect to the mean value of z that is 2, we have for the normalized power spectrum ρ:

$$G(\rho) = 1 - e^{-\rho} \tag{3}$$

Thus for a set of N observed values of random variables having a distribution similar to z, the joint distribution is given by:

$$G(\rho) = (1 - e^{-\rho})^N \tag{4}$$

This leads to the following statement:

Statement 1. *For a frequency range containing N power bins the probability that at least one bin has power greater than ρ times the mean of the noise power is given by:*

$$\mathcal{P}_{\mathcal{N}}(\text{power} > \rho) \approx Ne^{-\rho}, \text{ when } e^{-\rho} \ll 1 \tag{5}$$

The Eq. (5) allows us to determine a significance level (value of ρ) when $\mathcal{P}_{\mathcal{N}}$ is small.

2.1.2. Statistics of a sine wave embedded in a noise

The power spectrum of a pure sine wave embedded in a signal noise is defined as the sum of the Fourier spectrum components of the noise and of the deterministic signal and can be described as:

$$P(\nu_k) = \begin{cases} \left(x_r(\nu_0) + \frac{A}{2}\right)^2 + y_i^2(\nu_0) & , \nu_k = \nu_0 \\ x_r^2(\nu_k) + y_i^2(\nu_k) & , \nu_k \neq \nu_0 \end{cases} \tag{6}$$

Here the detail of the phase of the sine wave has been discarded because we are only interested in the power spectrum. At $\nu_k = \nu_0$, the power spectrum has no longer a χ^2 statistics with 2 d.o.f.; Instead it has a non-central χ^2 statistical distribution. Thus, if $z = (x + \frac{A}{2})^2 + y^2$ is the observed value in power spectrum at $\nu_k = \nu_0$ then it can be derived that its probability distribution is given by:

$$F(\rho, A) = \frac{e^{-\frac{A^2}{8}}}{2\pi} \int_0^{\rho} \int_0^{2\pi} e^{-\left(u - A\sqrt{\frac{u}{2}}\cos\theta\right)} \mathrm{d}\theta \mathrm{d}u \tag{7}$$

As reported by Gabriel et al. (2002). If we introduce the signal-to-noise ratio r as observed in the power spectrum as:

$$r = \frac{A^2}{8} \tag{8}$$

We can rewrite Eq (7) as:

$$F(\rho, r) = \frac{e^{-r}}{2\pi} \int_0^{\rho} \int_0^{2\pi} e^{-\left(u - 2\sqrt{r}\sqrt{u}\cos\theta\right)} \mathrm{d}\theta \mathrm{d}u \tag{9}$$

It's clear that when $A = 0$ the Eq. (9) is equal to Eq. (3).

Statement 2. *The probability that a sine wave with a deterministic amplitude A, having a signal-to-noise ratio r in the power spectrum ($r = A^2/8$), will have an observed power higher than a power level, ρ, is given by:*

$$\mathcal{P}_{\mathcal{A}}(\text{power} > \rho) = 1 - F(\rho, r) \tag{10}$$

2.2. Detection of long-lived modes embedded in short-lived modes

The detection of pure sine waves embedded in a spectrum of stochastically excited p modes can be done by testing the same hypotheses. Using the following technique we can reduce this particular case to the case of a sine wave embedded in a noise with χ^2 distribution:

1. Fit the p modes spectrum
2. Divide the power spectrum by the fitted model
3. Apply the statistical tests described in section 2.1

2.2.1. Statistics of a p-mode spectrum

The power spectrum of stochastically excited p modes can be described as:

$$P(\nu) = M(\nu_k)X(\nu_k) \tag{11}$$

Where $X(\nu_k)$ is a random function with a χ^2 statistical distribution with 2 d.o.f, and $M(\nu_k)$ is the model of the fitted mode made of a single Lorentzian profile plus noise.

One can fit a Lorentz profile to the observed power spectrum using Maximum Likelihood Estimators (MLE) technique (Toutain and Appourchaux, 1994), this is a classic and well-known approach used for the short-lived modes such as those observed on the Sun. After having done the fitting, one can divide the power spectrum by the fitted profile $\tilde{M}(\nu_k)$:

$$\tilde{X}(\nu_k) = \frac{P(\nu_k)}{\tilde{M}(\nu_k)} \approx X(\nu_k) \tag{12}$$

In first approximation $\tilde{X}(\nu_k)$ has also a χ^2 statistical distribution with 2 d.o.f. Thus the *statement 1* is applicable to $\tilde{X}(\nu_k)$, it is possible to apply the H0 statistical test to the corrected $\tilde{X}$ spectrum.

2.2.2. Statistics of a sine wave embedded in a p-mode spectrum

It can be derived that the power spectrum of a sine wave embedded in a p-mode spectrum can be described as:

$$P(\nu_k) = \begin{cases} M(\nu_0)\left(\left(x_c(\nu_0)+\frac{a}{2}\right)^2 + y_s^2(\nu_0)\right) & , \nu_k = \nu_0 \\ M(\nu_k)\left(x_r^2(\nu_k) + y_i^2(\nu_k)\right) & , \nu_k \neq \nu_0 \end{cases} \tag{13}$$

Where $a = \frac{A}{\sqrt{M(\nu_0)}}$

Dividing Eq.13 by $\tilde{M}(\nu_0)$ it becomes:

$$\tilde{X}(\nu_0) = \frac{M(\nu_0)}{\tilde{M}(\nu_0)}\left((x_c(\nu_0)+\frac{a}{2})^2 + y_s^2(\nu_0)\right) \tag{14}$$

Thus $\tilde{X}(\nu_0)$ can also be approximated to a non-central χ^2 statistical distribution. It is then possible to apply the H1 statistical test to the corrected power spectrum and to use the detection probability. expression presented in the *statement 2*.

Because the fitted p modes model ($\tilde{X}(\nu_k)$) is derived from data, it also has a statistical distribution that should be taken into account. Therefore we decided to use Monte-Carlo simulations for asserting the following:

1. The statistical distribution of the corrected power spectrum ($\tilde{X}(\nu_k)$) is close to a χ^2 with 2 d.o.f distribution.
2. The statistical distribution of an embedded sine wave signal in the corrected power spectrum is close to a non-central χ^2 statistical distribution with non-centrality parameter $\frac{A'^2}{4}$ (Statement 2 can be applied).

3. Validation of the detection technique

3.1. Monte-Carlo Analysis

In order to test our approximations and the H1 hypothesis, we performed a Monte-Carlo analysis of the aforementioned cases. We generated synthetic power spectra with a frequency resolution of 0.077μHz equivalent to an observation time length of 150 days. For both cases, we wanted to compare the analytical formula given by Eq (10) with detection levels as returned by the Monte-Carlo analysis.

3.1.1. Sine waves embedded in a noise signal

We repeated for 1000 simulations the following steps for a 30-μHz window:

1. Generate Fourier spectra for white noise
2. Add in the spectra three sine waves with different amplitudes as in Eq. (6)
3. Use MLE to estimate the mean value of the noise signal, σ
4. Determination for the power level: 10% probability in a 30-μHz gives $\rho = 8.26$ (Eq. 5)
5. Frequency extraction of the detected power bins

The power spectrum of one simulation is shown on Fig 1. After performing the simulations, we studied the distribution of the extracted frequencies. We counted the number of times that we detect each input sine waves signal and we compared the result with the predicted detection probability given by the Eq.9.

The significance level (Eq.5) depends on a proper estimation of the mean value of the noise. The implicit hypotheses that the set of power bins to be tested should be i.i.d random variables (See Section 2.1.1) does not hold in the presence of sine waves signal (Eq. (6)). The mean value of a non-central χ^2 distribution with a non-centrality parameter λ is given by $\lambda + \sigma$, where σ is the mean value of the central χ^2 distribution. In this context σ corresponds to the mean value of the noise signal, and λ corresponds to the observed amplitude in the power spectrum that the sine wave would have had if not embedded in the noise, i.e. $\frac{A^2}{4}$. This means that the inclusion in the noise signal data set of power bins with such statistical distribution

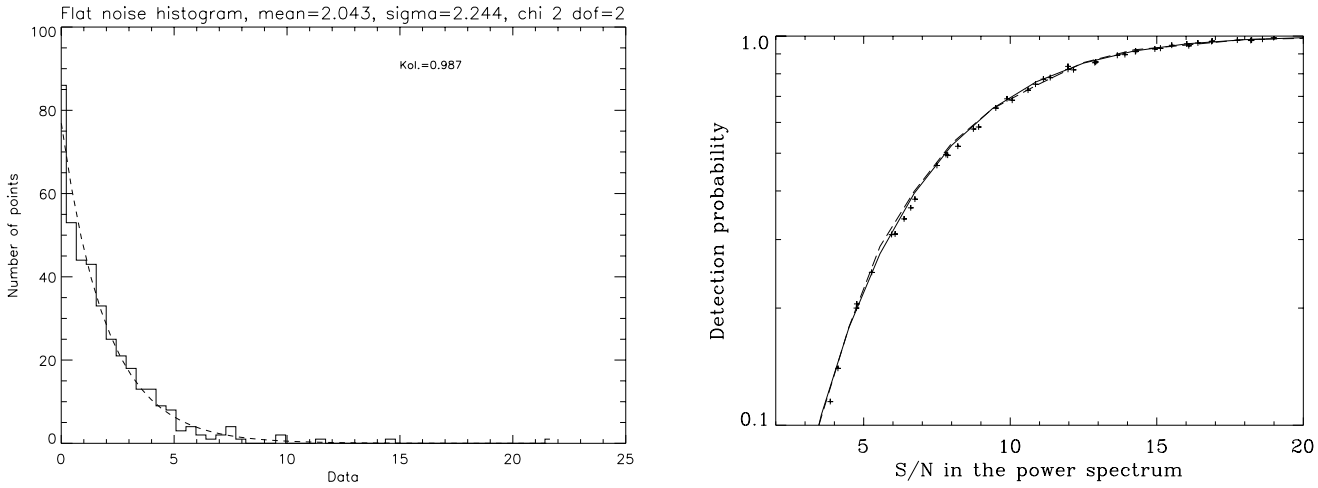

Fig. 2. Statistical analysis of the power spectra simulation of sine waves embedded in a white noise. On the left hand side: statistical distribution of the simulation shown in Fig.1. The statistical distribution of the power spectrum follows a χ^2 with 2 d.o.f statistics. On the right hand side: the agreement between the Eq.10 and the Monte-Carlo analysis. The solid line represents the detection probability as a function of the signal-to-noise ratio given by Eq.10. The dashed line represents the Monte-Carlo simulations result for one embedded sine wave. The measured detection ratios in 1000 Monte-Carlo simulations for different as a function of the signal-to-noise ratios are presented as a plus sign for the three embedded sine waves signal.

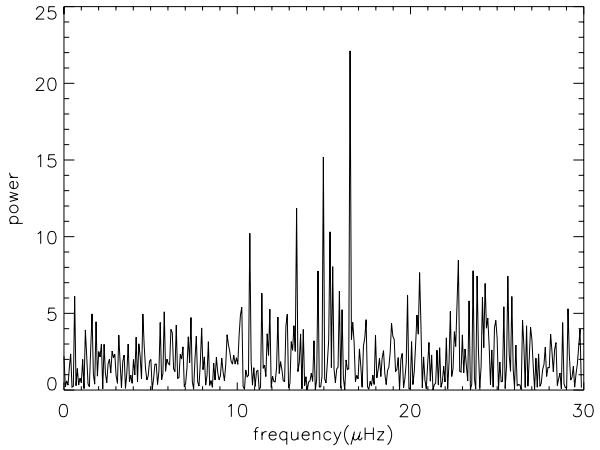

Fig. 1. An example of a Monte-Carlo simulation of sine waves embedded in a white noise. The signal due to the embedded sine waves is present at $\nu_k = 13.5, 15, 16.5$ (μHz). The input signal-to-noise ratios for the sine waves signal were 9, 10, 11, which correspond, respectively, to a theoretical 60%, 70%, 76% detection probability.

will produce an overestimation of the mean value of the noise signal (See Appendix A).

To avoid this problem, we first applied the statistical test H0 (Step 4) for the detection of outliers in the data set and then we re-estimated the mean value of the noise signal excluding the detected outliers. The value estimated with this additional step provided a less biased estimate of the true mean than even using the median.

The Fig.2 summarizes the results provided by our Monte-Carlo analysis for white noise only and the detection result as derived from *S*tatement 2. The similarity between statistical distributions can be measured by the Kolmogorov-Smirnov test. It computes the significance level (k) of the maximum absolute difference between two cumulative distribution functions. The computed k has values between 0, the smallest value of agreement, and 1, the largest value of agreement (see Press et al. 1988). We used the K-S to prove that the first approach could actually be done, we measured in 1000 simulations an averaged value $k = 0.8$.

3.1.2. Sine waves embedded in a p-mode spectrum

We repeated for 1000 simulations the following steps for a 30-μHz window:

1. Generate Fourier spectra for white noise
2. Generate a model of two short-lived p modes
3. Add in the spectra three sine waves with different amplitudes as in Eq. (13)
4. Use MLE for fitting the two p modes, and then divide the spectrum by the fit
5. Determination for the power level: 10% probability in a 30-μHz gives $\rho = 8.26$ (Eq. (5))
6. Frequency extraction of the detected power bins

For each simulation we applied the technique described in Section 2.2 (Step 4) and then followed the same procedure as for the sine waves embedded in pure noise signal described in the previous section (3.1.1). One example of the generated power spectrum and the respective corrected power spectrum are showed in Fig. 3.

The agreement between the Eq.(9) and the detection ratios of the input embedded sine waves signal in the 1000 simulations proves the second approach (Fig.3).

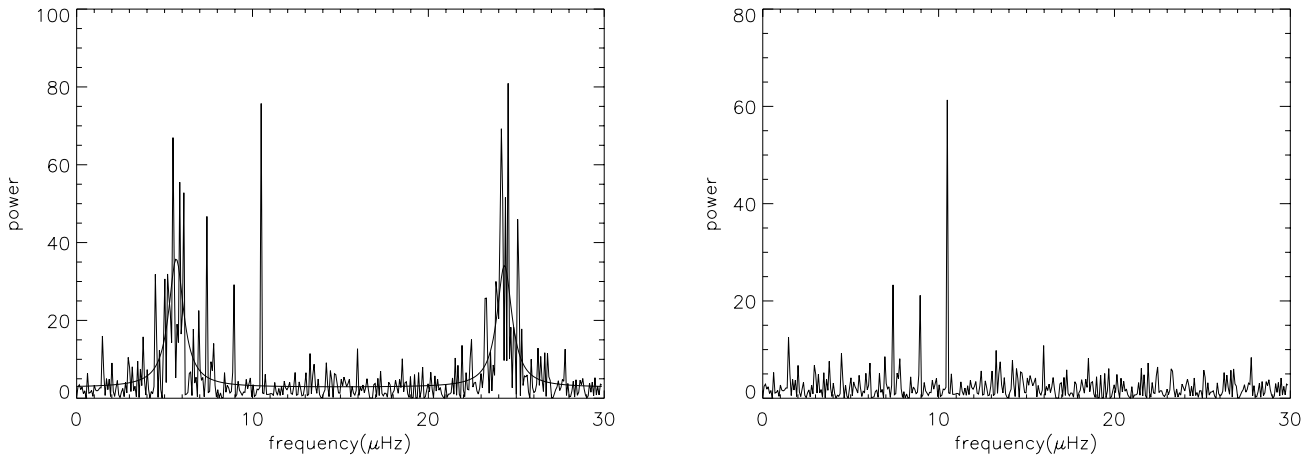

Fig. 3. On the left hand side: one example of a Monte-Carlo simulation of a sine wave embedded in a p mode. The signal due to the embedded sine waves is present at $\nu_k = 7.5, 9, 10.5$ (μHz). The solid line corresponds to the fitted model using the MLE. On the left hand side: The corrected power spectrum. The signal-to-noise ratios of the sine waves signal in the corrected power spectrum were input to be 9, 10, 11, which correspond, respectively, to a 60%, 70%, 76% detection probability.

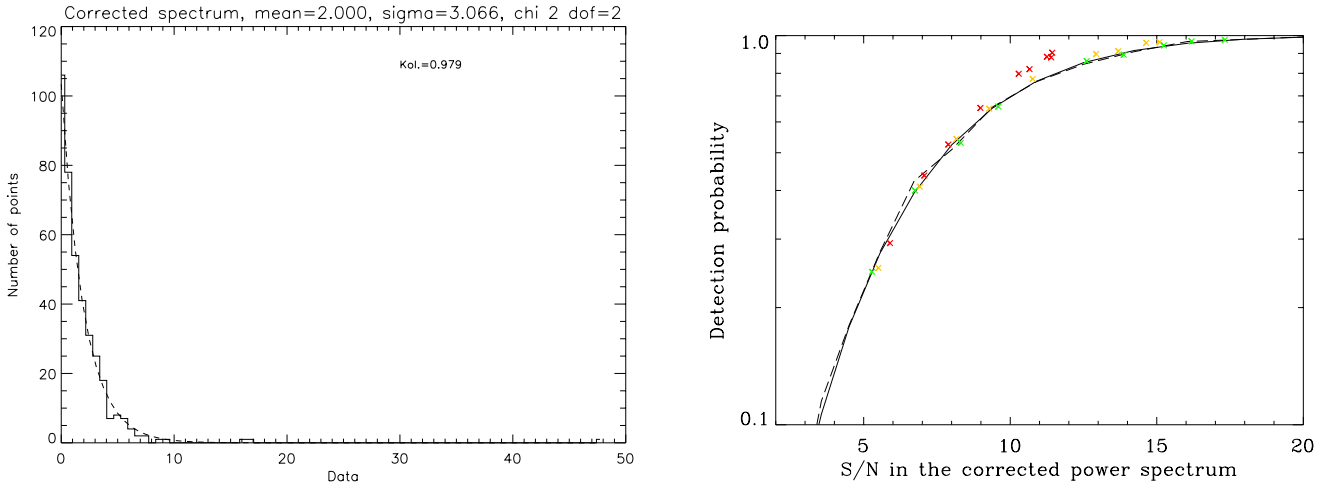

Fig. 4. Statistical analysis of the corrected power spectra simulation of sine waves embedded with stochastically excited p modes. On the left hand side: Statistical distribution of the corrected power spectrum for the simulation showed in Fig.3. It shows that the distribution follows a χ^2 with 2 d.o.f statistics. On the right hand side: The agreement between the Eq. (10) and the Monte-Carlo analysis. The solid line represents the detection probability as a function of the signal-to-noise ratio in the corrected power spectrum given by Eq. (10). The dashed line represents the Monte-Carlo simulations result for one embedded sine wave. The measured detection ratios in 1000 Monte-Carlo simulations for different as a function of the signal-to-noise ratios are presented as a plus sign for the three embedded sine waves signal.

3.2. Hare-and-Hound exercises of COROT: HD57006

3.2.1. The hares

A model of the star HD57006 has been computed with the CESAM code. It is a 1.65 $M_\odot$ model with a surface metallicity and a position in the HR diagram in agreement with the observations. A set of adiabatic eigenfrequencies and their normalized inertia for degrees $l = 1$ to 3 have been computed by one of the author (GB). Since the star has evolved beyond the main sequence stage, the frequency spectrum includes both p, g and mixed modes in the same frequency domain (Appourchaux et al, 2003). The time series of HD57006 was then generated by another author (TT) using the model frequencies computed by GB. The background noise was derived from the solar noise as measured by the SPM instrument aboard the SOHO spacecraft. The linewidth of the p modes was of the order of a 1 μHz, while that of the g modes was assumed to be much longer than the observation time.

3.2.2. The hounds

The first step in the analysis of the time series is to compute a power spectrum (Fig 5, left hand side), and then to construct and echelle diagramme (Fig 5, right hand side).

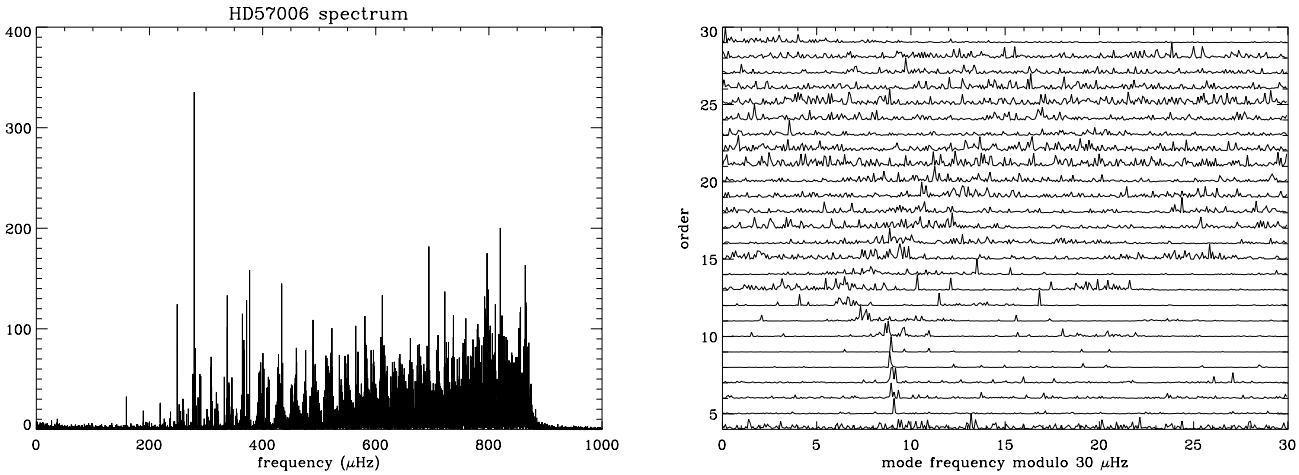

Fig. 5. On the left hand side is the spectrum of the star HD57006, and on the right hand side the Echelle diagram with $\Delta\nu_0$=30μHz.

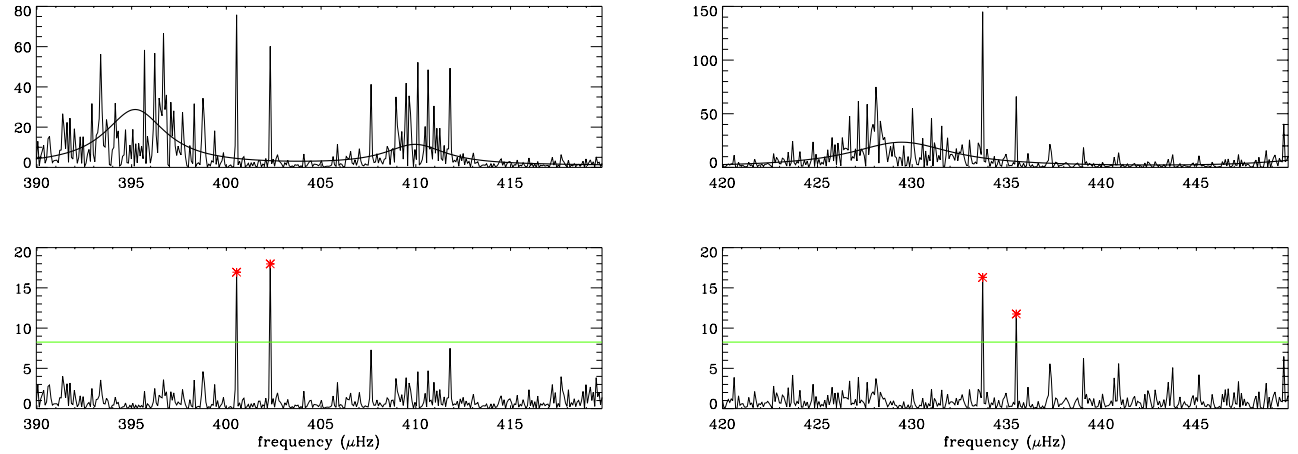

Fig. 6. Result of the test for two different ranges of frequencies. The fitting is presented on the top of the figure, and the division of the power spectrum by the fitting, as well as the bins above the 10% probability level, are presented on the bottom of the figure.

Echelle diagrammes are used for easing the mode identification as shown by Appourchaux (2003). One single ridge can clearly be identified as being that of the $l = 0$. The methodology described in section 3.1.2 was applied to 30-μHz windows between 200 and 500 μHz. Figure 6 shows two examples allowing to retrieve 4 long-lived modes.

3.2.3. Post-exercise analysis

The advantage of the HH exercise is the possibility of understanding what was simulated and how it was simulated. This is clearly not possible with *r*eal stars. Table 1 summarizes the narrow modes detected by our method together with the *a posteriori* mode identification, and the measured and theoretical signal-to-noise ratio. The probability are derived from Eq. (10) is also included in the table. One of the author (TT) provided us (TA and OM) with the theoretical signal-to-noise ratio for computing the number of detectable modes.

The number of detectable modes is simply the sum of the probability of detecting the narrow peaks as computed according to Eq. (10) (See Appendix B). The total number of detectable modes was 77 $\pm$ 9. This is higher by 4 σ than the 37 modes listed in Table 1. We would like to point out that the sampling effect of the Fourier transform can reduce the amplitude of the mode by up to factor 2 (Gabriel et al, 2002). This additional fact will also reduce the number of detectable modes but is not sufficient to explain the discrepancy which is related to the assumption made about the very nature of the g modes. We assumed that the modes have a long lifetime and behave like a pure sine wave of fixed amplitude. As a matter of fact, the hare (TT) generated a long-lived mode by stochastic excitation, i.e. using a multiplicative noise and not an additive noise. This has an impact on the detection probability that is indeed lower than for the case of additive noise (Eq (10)).

ν in μHz	Degree l	Azimuthal order m	$n_p - n_g$	Measured snr	Theoretical snr	Detection probability
226.08	1	0	-6	9.8	13.9	0.55
236.19	1	-1	-5	10.2	6.7	0.29
237.19	1	0	-5	16.8	11.9	0.50
252.39	1	-1	0	9.8	19.6	0.66
253.86	1	0	0	15.9	26.5	0.73
255.09	3	+1	-19	11.4	7.1	0.31
259.26	1	-1	-3	9.7	12.8	0.52
260.49	1	0	-3	12.2	24.1	0.71
265.59	2	-1	-10	13.0	3.8	0.11
268.60	2	+1	-10	7.5	5.3	0.21
276.62	2	+1	-9	37.1	19.9	0.66
278.32	2	+2	-9	7.3	6.1	0.26
279.78	1	-1	+2	41.0	22.9	0.70
281.10	1	0	+2	55.0	26.8	0.73
282.25	1	+1	+2	15.4	9.0	0.40
282.41	3	-1	-15	15.4	9.4	0.41
285.88	3	+1	-15	21.0	18.9	0.65
287.65	3	+2	-15	11.5	20.4	0.67
289.20	1	0	-1	36.9	60.7	0.87
290.66	1	+1	-1	35.5	40.7	0.81
296.68	2	-1	-7	7.3	10.7	0.45
301.70	2	-2	-6	7.8	6.7	0.29
303.40	2	-1	-6	9.7	12.5	0.52
311.11	3	-2	-11	22.8	6.6	0.29
315.82	2	-1	-5	7.9	13.4	0.54
318.21	3	+2	-11	11.1	18.6	0.64
332.25	2	-1	-4	14.4	9.7	0.43
345.76	3	+1	-8	14.2	19.5	0.65
347.53	3	+2	-8	7.3	24.1	0.71
363.12	2	-1	-1	8.1	18.5	0.64
364.27	2	+1	-2	29.7	12.5	0.52
371.68	3	-1	-8	14.23	7.6	0.34
377.01	3	+2	-8	33.2	13.0	0.53
400.54	3	-2	-3	10.4	13.5	0.54
402.31	3	-1	-3	8.0	23.4	0.70
433.72	3	-1	-1	20.2	4.9	0.19
435.49	3	0	-1	8.6	3.1	0.07

Table 1. Frequencies of the g modes extracted using the statistical test described in the section 2.2. The identification in terms of l, m and $n_p - n_g$ was performed *a posteriori*. The measured signal-to-noise ratio (snr) can be compared to that of the theoretical. The detection probability is computed after Eq. (15) using the theoretical snr.

In the case of multiplicative noise, Eq. (10) is then replaced by the following:

$$\mathcal{P}_{\mathcal{A}}(\text{power} > \rho) = 1 - G(\rho/(r+1)) = e^{-\frac{\rho}{r+1}} \qquad (15)$$

For a signal-to-noise ratio greater than 10 the probability of detecting a mode is a factor two third lower for a multiplicative noise (Eq. (15)) than for an additive noise (Eq. (10)). Taking into account Eq (15), the number of modes becomes then 48 ± 7. That is about 1.5 σ away from the number of modes effectively detected, thereby confirming Eq. (15). In summary, it is more difficult to detect a long-lived mode that is stochastically excited than to detect a pure sine wave. This finding is the result of the *a posteriori* analysis and is not strictly part of the hare-and-hound exercise.

4. Conclusion

We showed that we could detect long-lived modes embedded into short-live p modes. This occurs for evolved solar-like stars having long-lived mixed modes. These modes appear due to the increase of the Brunt-Vaisäla frequency during the Helium core evolution phase. The detection of such mixed modes will provide a powerful tool for understanding the internal evolution of the stars. Here we introduce a new technique that allows, after having fitted the short-lived p modes, to detect these long-lived mixed modes. The steps of the detection technique are summarized in Section 3.1.2. The technique was first used in the frame work of the COROT hare-and-hound exercise. The reliability of this technique was then tested and validated by performing Monte-Carlo simulations. We showed during the hare-and-hound-exercise post analysis, that long-

lived modes excited stochastically are more difficult to detect that deterministic sine waves.

Acknowledgements. Olga Moreira was supported by a grant from the portuguese Agência de Inovação.

References

Andersen, B.N. 1996, A&A, 312, 610
Appourchaux, T. 2003, Ap&SS, 284, 109
Appourchaux, T., Moreira, O., Berthomieu, G., & Toutain, T. 2003. In "Stellar structure and habitable planet finding", 2nd Eddington workshop, F. Favata, S.Aigrain eds., ESA-SP,538
Appourchaux, T., Fröhlich, C., Andersen, B., et al. 2000, ApJ, 538, 401
Gabriel, A.H., Baudin, P., Garcia, R.A., et al. 2002, A&A, 390, 1119
Grec, G. 1981, Ph.D. thesis, Université de Nice
Kumar, P., Quataert, E.J., & Bahcall, J.N. 1996, ApJ, 458, L83
Morel, P. 1997, A&A Sup. Series, 124, 597
Press,W. H., Flannery, B. P. , S. A. Teukolsky, S. A., & Vetterling, W. T. 1988, Numerical recipes in C, Cambridge Univ. Press.
Tassoul, M. 1980, ApJS, 43, 469
Toutain, T., & Appourchaux, T. 1994, A&A, 289, 649
Turck-Chièze, S., García, R., Couvidat, S., 2004, ApJ, in press

Appendix A: MLE of the mean noise value

Considering a noise signal with a χ^2 with 2 d.o.f statistics, and σ' the MLE of the mean, σ, then:

$$L = \prod_{i=1}^{N} \frac{1}{\sigma'} e^{-\frac{z_i}{\sigma'}} \tag{A.1}$$

or:

$$l = -\sum_{i=1}^{N} \left(\sigma' + \frac{z_i}{\sigma'}\right) \tag{A.2}$$

minimization of l:

$$\frac{dl}{d\sigma'} = \frac{N}{\sigma'^2}\left(\frac{1}{N}\sum_{i=1}^{N} z_i - \sigma'\right) = \frac{N}{\sigma'^2}\left(\overline{z} - \sigma'\right) \tag{A.3}$$

So $\sigma' = \overline{z}$ is an unbiased estimator of σ, $(E[\sigma'] = \sigma)$, with variance $\frac{\sigma'^2}{N}$.

When there is an embedded signal with a different statistics not all z_i are independent and identically distributed (i.i.d) random variables and $\sigma' = \overline{z}$ is no longer an unbiased estimator of σ. If z'_i are i.i.d variables with χ^2 with 2 d.o.f distribution with mean σ and z''_j variables with a non-central χ^2 statistics with non-centrality parameters λ_i , then for $Z = Z' + Z''$:

$$E[\overline{z}] = E\left[\frac{1}{N}\left(\sum_{i=1}^{N-p} z'_i + \sum_{j=1}^{p} z''_j\right)\right] \tag{A.4}$$

$$= \frac{N-p}{N}E[z'] + \frac{1}{N}\sum_{j=1}^{p} E[z''] \tag{A.5}$$

$$= \frac{N-p}{N}\sigma + \frac{1}{N}\sum_{j=1}^{p}(\lambda_i + \sigma) \tag{A.6}$$

$$= \sigma + \frac{\sum_{j=1}^{p}\lambda_i}{N} \tag{A.7}$$

Thus the bias, B, is given by:

$$B(\sigma') = E[\sigma'] - \sigma = \frac{\sum_{j=1}^{p}\lambda_i}{N} \tag{A.8}$$

Which means that $E[\overline{z}] \approx \sigma$ when $\frac{\sum_{j=1}^{p}\lambda_i}{N} \ll 1$.

Appendix B: Average number of detectable peaks amongst N peaks

A proper analysis of the number of detectable modes should take into account the probability of detecting each long-lived mode as generated by the hare. The mean number of detectable modes can be derived from the property of the N Bernouilli random variable X_j, each having a detection probability p_j for which we have:

$$E[X_j] = p_j \tag{B.1}$$

and the standard deviation is given by:

$$\sigma_{X_j} = p_j(1 - p_j) \tag{B.2}$$

In our case, the observed random variable is the sum of the N random variable X_j written as:

$$X = \sum_{i=1}^{i=N} X_j \tag{B.3}$$

Since the N random variable are independent from each other, we have:

$$E[X] = \sum_{j=1}^{j=N} p_j \tag{B.4}$$

The standard deviation is then given by:

$$\sigma_X = \sqrt{\sum_{j=1}^{j=N} p_j(1 - p_j)} \tag{B.5}$$